A time-efficient re-calibration algorithm for improved long-term accuracy of head-worn eye trackers

Christian Lander* DFKI GmbH
Frederic Kerber* DFKI GmbH
Thorsten Rauber† Saarland University
Antonio Krüger* DFKI GmbH

Abstract

Mobile gaze-based interaction has been emerging over the last two decades. Head-mounted eye trackers as well as remote systems are used to determine people's gaze (e.g., on a display). However, most state-of-the-art systems need calibration prior to usage. When using a head-mounted eye tracker, many factors (e.g., changes of eye physiology) can influence the stability of the calibration leading to less accuracy over time. Re-calibrating the system at certain time intervals is cumbersome and time-consuming. We investigate methods to minimize the time needed and optimize the process. In a user study with 16 participants, we compared partial re-calibrations with different numbers of calibration points and types of adaptation strategies. In contrast to a full calibration with nine points, the results show that a re-calibration with only three points results in 60% less time needed and achieves a similar accuracy.

Keywords: eye tracking, re-calibration, gaze estimation

Concepts: •**Computing methodologies** → *Computer vision;*

1 Introduction

Gaze has been studied as an input technique for over 25 years and is known to be faster than other pointing devices (e.g., mice [Sibert and Jacob 2000]). With gaze, people usually indicate what is attracting them and might be of interest [Vertegaal 2003]. Common applications include desktop control [Jacob 1990], eye-typing [Majaranta and Räihä 2002] and target selection [Stellmach and Dachselt 2012]. Advances made in head-mounted eye tracking point toward pervasive eye gaze interfaces for everyday usage [Bulling and Gellersen 2010].

Head-mounted eye trackers based on the Pupil Center Corneal Reflection (PCCR) technique are usually equipped with two cameras: (1) an eye camera capturing a close-up view of a person's pupil position, and (2) a scene camera partly recording the field of view. A user-dependent calibration [Duchowski 2007] is needed to create a mapping from the 2D pupil positions in the eye to the 2D gaze positions in the scene camera's coordinate system. Although research has been done to enhance the calibration procedure, this still remains the major problem of eye tracking systems [Pfeuffer et al. 2013]. Head-mounted eye trackers have to be repeatedly re-calibrated, due to calibration drift. Reasons for this include changes in the eye physiology (e.g., wetness of the eye), the environment (e.g., lighting conditions) and orientation of and/or distance to the target display. Hence, the use of state-of-the-art head-mounted eye tracking systems is cumbersome in prolonged interaction scenarios. Existing re-calibration methods use one on-screen re-calibration point [Stampe 1993; Stampe and Reingold 1995] or require additional input methods [Hornof and Halverson 2002].

In this paper, we present a new approach that periodically invokes a partial re-calibration procedure to counteract the calibration drift in real time. Our method takes far less time than the standard calibration routine while achieving competitive gaze estimation results. We use a subset of the full calibration points for re-calibration. The information gathered during this process is used to establish an updated mapping function with new coefficients that will improve gaze estimation accuracy after a calibration drift.

In a laboratory experiment with 16 participants, we evaluated our re-calibration approach against a nine-point full calibration while simulating two different sources of calibration drift, i.e. changing the distance to the target display as well as putting on and taking off the device. We tested four different adaption methods and analyzed how many re-calibration points and which combinations of their distributions deliver the best results. We found that a proper selection of calibration points is far more important than the selection of the adaption method. Thus, our contributions are twofold:

- **Fast re-calibration** of a head-mounted eye tracker improving the gaze estimation accuracy after a calibration drift.
- **Optimal re-calibration point combinations and adaption methods** leading to the best possible improvement of gaze estimation accuracy on a display.

In addition, we define two rules that make it possible to derive the optimal calibration-point selection.

2 Time-Efficient Re-Calibration

Estimating a user's gaze using a head-mounted eye tracker faces one major challenge: the eye tracker has to be repeatedly re-calibrated due to calibration drift, decreasing the current gaze estimation accuracy. In the following, we will describe the standard calibration procedure and our new time-efficient approach. Both are based on the open-source head-mounted eye tracking framework Pupil [Kassner et al. 2014] and the corresponding device.

2.1 Eye Tracker Calibration

Head-mounted eye trackers are calibrated, prior to usage, to map 2D pupil positions from the eye to 2D gaze positions in the scene camera's coordinate system. In the case of a nine-point calibration, data is sampled as follows: Facing the target display covered by the scene camera's field of view, the user fixates a visual marker shown at nine different on-screen locations $c \in C$ in a row, each for a certain time interval. During this time span, the system records calibration point tuples *cp*, containing the 2D positions of the pupil $(x_{1_{cal}}, y_{1_{cal}})$ and the corresponding visual marker $(x_{2_{cal}}, y_{2_{cal}})$ in their respective camera coordinate systems (Figure 1).

*e-mail: firstname.lastname@dfki.de
†e-mail: s9thraub@stud.uni-saarland.de

ETRA 2016, March 14 - 17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03...$15.00
DOI: http://dx.doi.org/10.1145/2857491.2857513

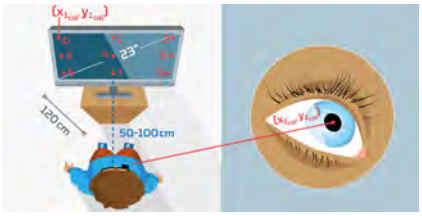

Figure 1: *Illustration of pupil and calibration point coordinates, as well as study setup, calibration points 0-8.*

The set of calibration points Cal is used to calculate the coefficients for the mapping function M_{Cal} between the two camera coordinate systems. The coefficients are obtained by solving a linear equation system – one equation per calibration point – aggregating the points in Cal accordingly. Once calculated, M_{Cal} can be changed by modifying the coefficients leading to different gaze estimations.

2.2 Time-Efficient Re-Calibration

Our approach uses a subset of the initial calibration points to reduce the time spent for re-calibration. This minimizes interruption and distraction for the user while preserving the system's accuracy, even for longer-lasting eye tracking sessions. The set of re-calibration points $R, R \subset C$ delivers new re-calibration tuples rp consisting of pupil and on-screen marker positions for every location $r \in R$:

$$rp_{Recal} = (x_{1_{recal}}, y_{1_{recal}}, x_{2_{recal}}, y_{2_{recal}}).$$

With the set of re-calibration points *Recal*, we compute the offset between the values Cal_{cp}, $cp \in Cal$ recorded during the full calibration and current re-calibration values $Recal_{rp}$ for every re-calibration point $rp \in Recal$:

$$\begin{aligned} \textit{offset}_{rp} &= Cal_{cp} - Recal_{rp} \\ &= (x_{1_{cal}}, y_{1_{cal}}, x_{2_{cal}}, y_{2_{cal}}) - (x_{1_{recal}}, y_{1_{recal}}, x_{2_{recal}}, y_{2_{recal}}). \end{aligned}$$

The computed offset values $\textit{offset}_{rp}$ are used to estimate the changes for the values Cal_{cp} of those calibration locations c that were not part of the re-calibration, $c \in (C \setminus R)$. The estimated changes lead to updated values $Cal'_{cp}, \forall cp \in (Cal \setminus Recal)$ that can be used to create an updated mapping function M_{Recal} in combination with the values $Recal_{rp}, \forall rp \in Recal$. In the case that $p \in R$, the values of all calibration point recordings are replaced by the ones obtained during re-calibration; otherwise they are updated based on the shift that was recognized. Hence, the updated mapping function M_{Recal} yields more accurate results than the original mapping function M_{Cal}.

We used different methods to estimate the changes in pupil and calibration-marker positions of those calibration points that were not part of the re-calibration:

- Overall Offset (*OO*) computes a tuple o_{cp} out of the re-calibration points' offset values $\textit{offset}_{rp}$:

  $$o_{cp} = \frac{\sum_{r \in R} \textit{offset}_{rp_k}}{|R|}, with\ k \in x_1, y_1, x_2, y_2.$$

  This tuple o_{cp} is added to all tuples Cal_{cp} where $c \in (C \setminus R)$, to obtain the updated calibration values $Cal'_{cp} = Cal_{cp} + o_{cp}$.

- Averaged Offset of Nearest Neighbors (*NN*) computes a distinct offset for every point $c \in (C \setminus R)$ based on the averaged offsets of the point's nearest neighbors that were part of the re-calibration. The sets of nearest neighbors

  $$NN_{cp} = \{p \in R \mid d(cp, p)\ is\ minimal\},$$

  are calculated using the Euclidean distance w.r.t. the ratio of the calibration area (NNR) or normalized values (NNN) distance function. Finally, the averaged offset o_{cp},

  $$o_{cp} = \frac{\sum_{p \in NN_{cp}} \textit{offset}_{rp_k}}{|NN_{cp}|}, with\ k \in \{x_1, y_1, x_2, y_2\},$$

  is computed for every point $c \in (C \setminus R)$, to update the calibration values $Cal'_{cp} = Cal_{cp} + o_{cp}$.

- Inverse Distance Weighting (*IDW*) computes a distinct offset for every calibration location $c \in (C \setminus R)$ based on the inverse distance weighting [Shepard 1968]. The overall offset tuple o_{cp} is computed by multiplying a weighting *w(rp, cp, exp)* ($\forall r \in R$ *and* $c \in (C \setminus R), exp = 2$) with individual offset values $\textit{offset}_{rp}$:

  $$o_{cp} = \sum_{r \in R} w(rp, cp, exp) \times \textit{offset}_{rp_k}, with\ k \in \{x_1, y_1, x_2, y_2\}$$

  The computation for the updated values Cal'_{cp} remains the same as in the previous approaches.

2.3 Implementation

Our system consists of two components: (1) a monocular head-mounted Pupil eye tracker connected to a laptop; and (2) a 23-inch display. The eye tracking re-calibration algorithms are written in Python and integrated as a plugin in Pupil's open-source framework running in real time. Thus, the re-calibration methods can be accessed via a graphical user interface. To update and compute the mapping function, the scientific computing package NumPy[1] is used. For gaze estimation on the display, Pupil's integrated Marker Tracking plugin[2] is used to define and track surfaces. Therefore, a set of four visual markers on the display corners for tracking the orientation between the eye tracker and the display is used.

3 Evaluation

We conducted a laboratory user study to assess the effectiveness of our approach as well as the optimal adaption method and the set of re-calibration points. To simulate calibration drift, we (1) changed the distance to the target display (*50 cm* vs. *100 cm*) and (2) participants took off the eye tracker between measurements (*take-off* vs. *keep-on*). To test all combinations, we conducted several experiment blocks, each consisting of two runs (see Table 1). In runs 6 and 8, we only observed the calibration drift over time.

Each test run was divided into two parts: (1) the nine-point full calibration, which enabled us to evaluate all our re-calibration methods with all possible point combinations post-hoc; (2) a gaze pointing task to sample data for the analysis. Therefore, we divided the screen in a 6×4 grid. All participants had to fixate the center of each area, indicated in random order by a red dot, for three seconds.

In total, 16 test runs (8 blocks with two runs each) were recorded per participant. To reduce fatigue effects, we split the procedure into two sessions (45 minutes duration on average) and introduced a break of about 90 seconds between two subsequent test runs. We counterbalanced the execution order across participants using a Latin square. Data was sampled at 30 Hz (i.e., 90 samples per on-screen target). We discarded 1.5 seconds (first 35, last 10 samples) of the data, as pre-tests showed that this time was needed to look for the targets in the worst case. Based on the remaining 45 samples, the average fixation position was computed.

[1] www.numpy.org
[2] www.pupil-labs.com/blog/2013/12/036-release.html

Exp. block		1	2	3	4	5	6	7	8
Dist. in cm	run 1	50	50	100	100	50	50	100	100
	run 2	100	100	50	50	50	50	100	100
Device taken off		✓	✗	✓	✗	✓	✗	✓	✓

Table 1: *Design of the experiment blocks and test runs.*

The collected data consisted of the following: For each block, we got an initial mapping M_{First} for the first full calibration and another mapping M_{Second} for the subsequent full calibration. Corresponding to each of the mappings, we also got a set of 24 test recordings evenly distributed across the screen. Based on a subset R of the calibration points used in M_{Second} and the initial mapping M_{First}, we were able to compute mapping functions M_{Recal_R}, as presented above. These newly created mapping functions based on a subset of calibration points are those to be evaluated against the ground-truth function M_{Second}. For the evaluation itself, we therefore examined the test data that was recorded after the second full calibration and evaluated it with respect to average gaze estimation error in degrees of visual angle – first with the mapping function M_{Second} as baseline and then with the newly created mapping functions M_{Recal_R}.

The study setup was as follows (Figure 1): we used a 23" full HD monitor, placed on a table. To track the display in the scene camera's field of view and map gaze accordingly, we attached four markers at the corners and used the marker tracking plugin provided by the Pupil framework. We used two distances *50 cm* and *100 cm* to the display, and taking off the eye tracker between measurements (*take-off* vs. *keep-on*). In order to test all combinations of the two independent variables, we conducted several experiment blocks, each consisting of two test runs (see Table 1). Runs 6 and 8 represent the simplest cases as we did not simulate any changes.

16 participants (four female), between 19–58 years old (M=28.5 years, SD=10.44 years) with a body height between 160–186 cm (M=174.44 cm, SD = 8.13 cm) were recruited. All of them had normal or corrected-to-normal vision. To ensure optimal pupil tracking, they were asked to participate without makeup.

4 Results and Discussion

To assess the accuracy of the initial full calibrations, we calculated the average error in degrees of visual angle. This value describes the difference of visual angle between the computed gaze point and the actual fixation targets on the screen. We performed a paired-sampled t-test between the initial calibration M_{First} ($M = 1.18°, SD = 0.69°$) and the subsequent calibration M_{Second} ($M = 1.41°, SD = 1.07°$) across all participants and all study blocks. It revealed a significant difference for accuracy ($t(127) = 3.08, p < 0.05$), that mainly originates from fatigue effects of the participants. As participants had to fixate $8 \times 24 = 192$ targets and $8 \times 9 = 72$ calibration points, this result emphasizes the need to reduce the number of fixations required for (re-)calibrations.

We evaluated the second test data set by simply using the initial mapping function M_{First}, decreasing the accuracy to $4.91°$ as expected. A paired-samples t-test showed a significant difference between the initial calibration ($M = 4.91°, SD = 4.4°$) and the subsequent one ($M = 1.41°, SD = 1.07°$), $t(127) = 9.28, p < 0.05$.

The key questions for our approach are (1) how many and (2) which points should be used for a time-efficient re-calibration to improve the accuracy. We examined only re-calibrations with five points or less, resulting in time saving of 40–90% compared to a full calibration. For all other cases the average improvement in accuracy drops below $0.1°$. The analysis of our four different re-calibration methods revealed slight differences (below $0.1°$ on average) among them. A Greenhouse-Geisser corrected one-way ANOVA showed a significant difference between the four methods ($F(1.234, 156.708) = 7.989, p < 0.05$) for $|R| = 2$. A pairwise post-hoc analysis with Bonferroni correction in effect revealed significant differences ($p < 0.05$) between all combinations of re-calibration methods, but NNN and OO as well as IDW and OO.

In the experiment, we focused on two conditions that increase calibration drift, i.e. removing and replacing the device as well as changing the distance to the target display. Applying the mapping functions M_{First} and M_{Second} to the second test run, we found highest error ($M > 7°$) to occur when both the distance was changed and the eye tracker was taken off (blocks 1 and 3). In cases without any changes, we carried out the lowest additional error of $0.9°$ (blocks 6 and 8).

Figure 2 shows the differences in the gaze estimation errors between the mapping functions M_{Recal_R} (using $|R|$ re-calibration points) and M_{Second}. To highlight the improvement compared to the outdated mapping function M_{First}, we also added the difference between M_{First} and M_{Second}. We observed that our approach is highly effective in counteracting calibration drift caused by removing the device (blocks 5 and 7). On average, re-calibrations with one or two points decrease the additional gaze estimation error below $0.5°$. However, in cases in which the distance between user and target display is changed (blocks 1-4), more calibration points are required to achieve high accuracy values. Depending on the use case, a minimum of three points should be used in these cases.

We further analyzed the distribution of points that should be chosen for a re-calibration, as a proper selection might also further reduce the necessary point amount: In the case of $|R| = 1$, our four proposed adaption methods NNN, NNR, IDW and OO do not establish different mapping functions, resulting in nine different re-calibrations. A Greenhouse-Geisser corrected one-way ANOVA showed a significant difference in accuracy with respect to the re-calibrations and the standard full calibration M_{Second} ($F(2.664, 39.958) = 42.173, p < 0.05$). Pairwise Bonferroni corrected post-hoc tests revealed significant differences in all cases ($p < 0.05$) but in using point 6, showing the lowest error. This contradicts [Stampe 1993], suggesting the center point. This could be related to the use of a monocular device tracking the right eye, with the greatest distance to point 6.

For $|R| > 1$, we separately analyzed the four proposed adaption methods NNN, NNR, IDW and OO. We created two sets S_p and NS_p that contain those point selections with p points that result in significantly / insignificantly different accuracy w.r.t. the baseline M_{Second}. For the point selections in the sets NS_p, the null hypothesis that there is no difference between the means of gaze estimation error cannot be rejected. This means that we were able to create an

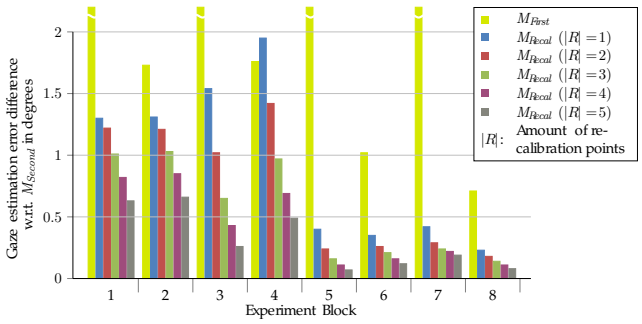

Figure 2: *Difference in the gaze estimation error w.r.t. M_{Second}.*

effective mapping function with the respective re-calibration combination in this special case. However, we cannot make general assumptions about the suitability of this created mapping function.

To further analyze the point distributions, we defined two rating functions to express the calibration area coverage achieved by the selected points. For the case of $|R| = 2$ we consider the covered axes intercepts based on the following formula:

$$axisIntercept((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$$

with normalized re-calibration coordinates (x_1, y_1) and (x_2, y_2). The highest value of 2 is achieved by selecting two corner points on the diagonal, whereas two adjacent points parallel to one of the axes yield the lowest value of 0.5. A Welch's unequal variances t-test showed a significant difference in the axis intercept values between the groups $NS_2(M = 1.33, SD = 0.49)$ and $S_2(M = 0.95, SD = 0.42); t(20.74) = 3.17, p < 0.05$.

For $|R| > 2$, we consider the area that is covered by the polygon of the selected calibration points. The area of an arbitrary polygon that is not self-crossing can be computed with the following formula:

$$area((x_1, y_1), ..., (x_n, y_n)) = 0.5 * (x_1y_2 - x_2y_1 + x_2y_3 \\ -x_3y_2 + ... + x_{n-1}y_n - x_ny_{n-1} + x_ny_1 - x_1y_n),$$

with normalized re-calibration coordinates $(x_1, y_1), ..., (x_n, y_n)$. Table 2 prints the Welch's unequal variances t-tests, that showed a significant difference in the covered area for all cases.

$\|R\|$	M(NS)	SD(NS)	M(S)	SD(S)	t-test ($p < 0.05$)
3	0.28	0.15	0.19	0.12	$t(97.06) = 4.49$
4	0.48	0.18	0.37	0.13	$t(273.61) = 7.31$
5	0.62	0.18	0.50	0.13	$t(436.62) = 8.60$

Table 2: *Results of Welch's unequal variances t-test of the covered polygon area between $NS_{|R|}$ and $S_{|R|}$.*

Based on the results from above, we recommend two rules for selecting the appropriate point distribution:

1. Axes intercept values / polygon area should be maximized.
2. Calibration points in the corners should be preferred.

Following these rules, the following sets of points should be chosen for re-calibration in our setting using a nine-point calibration:

#	Combinations
2	[0,8], [2,6]
3	[0,2,6], [0,2,8], [0,6,8], [2,6,8]
4	[0,2,6,8]
5	[0,1,2,6,8], [0,2,3,6,8], [0,2,4,6,8], [0,2,5,6,8], [0,2,6,7,8]

Table 3: *Point combinations yielding the best results, chosen based on the two rules defined above.*

In total, our results suggest that OO and IDW are more robust w.r.t. re-calibration point distribution. This is not surprising when considering the strategy behind the approaches. In the case of OO, all available re-calibration points are considered to compute the missing points, thereby providing the most global approach. When considering only combinations with a proper point distribution, the difference in gaze estimation error between the adaption methods drops below 0.05°. This clearly shows that a meaningful point distribution is much more important than the adaption method.

5 Conclusion and Outlook

The paper describes a method for a time-efficient re-calibration of head-mounted eye trackers after a calibration drift. In contrast to existing approaches, the method works in real time, without the need of further input devices, while improving the accuracy towards the original calibration. It can handle sudden increases of calibration drift and considers the varying error across the screen. Our experiment showed that our method compares well against repeating a full calibration with nine points. In addition, we recommended rules for the set and distribution of re-calibration points. With our approach, the use of head-worn eye trackers can be improved, especially in long-term settings that suffer from disruptive re-calibrations. In the future, we want to take our approach one step further and do an evaluation of our re-calibration approach in a real-world scenario (e.g., in final industrial assembly). Furthermore, we plan to contribute to the Pupil community by providing our re-calibration plugin to the open source framework.

References

BULLING, A., AND GELLERSEN, H. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing 9*, 4, 8–12.

DUCHOWSKI, A. T. 2007. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc.

HORNOF, A., AND HALVERSON, T. 2002. Cleaning Up Systematic Error in Eye-tracking Data by Using Required Fixation Locations. *Behavior Research Methods, Instruments, & Computers 34*, 4, 592–604.

JACOB, R. J. 1990. What You Look At is What You Get: Eye Movement-Based Interaction Techniques. In *Proc. CHI '90*, 11–18.

KASSNER, M., PATERA, W., AND BULLING, A. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-Based Interaction. In *Adj. Proc. UbiComp '14*, 1151–1160.

MAJARANTA, P., AND RÄIHÄ, K.-J. 2002. Twenty Years of Eye Typing: Systems and Design Issues. In *Proc. ETRA '02*, 15–22.

PFEUFFER, K., VIDAL, M., TURNER, J., BULLING, A., AND GELLERSEN, H. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proc. UIST '13*, 261–270.

SHEPARD, D. 1968. A Two-dimensional Interpolation Function for Irregularly-Spaced Data. In *Proc. ACM '68*, ACM '68, 517–524.

SIBERT, L. E., AND JACOB, R. J. 2000. Evaluation of Eye Gaze Interaction. In *Proc. CHI '00*, 281–288.

STAMPE, D. M., AND REINGOLD, E. M. 1995. Selection by Looking: A Novel Computer Interface and its Application to Psychological Research. *Studies in Visual Information Processing 6*, 467–478.

STAMPE, D. M. 1993. Heuristic Filtering and Reliable Calibration Methods for Video-Based Pupil-tracking Systems. *Behavior Research Methods, Instruments, & Computers 25*, 2, 137–142.

STELLMACH, S., AND DACHSELT, R. 2012. Look & Touch: Gaze-Supported Target Acquisition. In *Proc. CHI '12*, 2981–2990.

VERTEGAAL, R. 2003. Attentive User Interfaces. In *Communications of the ACM*. 30–33.